I/S: A JOURNAL OF LAW AND POLICY FOR THE INFORMATION SOCIETY

L. JEAN CAMP*

Reliable, Usable Signaling to Defeat Masquerade Attacks

Abstract: The great irony of our information infrastructure is that while there is an immense variety of data, it flows through a single channel. Therefore, individuals are forced to evaluate the safety of the web sites they are viewing on the same channel that may be controlled by a malicious party. How can a consumer obtain the information necessary to discriminate between good and bad resources? The solution to this particular signaling problem is to create mechanisms that are embedded in the communication but beyond the control of the web site. Net Trust is an application that informs the web-browsing experience with an individually tailored, easy-to-use, thoroughly tested interface. Net Trust is focused on authenticating web sites *to individual users* instead of authenticating individual users *to web sites*. This paper describes the theory, motivation, and interaction behind Net Trust.

* The author is an Associate Professor in the School of Informatics at Indiana University, Bloomington, Indiana.

I. INTRODUCTION

When on the Internet, individuals are susceptible to a vast number of risks, which include unreliable software, malicious web sites, and old-fashioned merchant fraud schemes disguised in new electronic trappings. There are both centralized and distributed mechanisms to assist individual users to detect malicious actors and protect themselves on the network. Empowering online communities by enabling peer production of security and privacy information is critical to helping individuals protect themselves online.

In this paper, I describe a system that not only allows individuals to easily access and annotate their own histories, but also to share those histories with a self-selected social network. Net Trust uses both socially generated information and centralized data to enable informed decision-making about the web sites users visit. By integrating both social networking and centralized information, the system provides both personal histories and centralized information sources to the individual user in one simple interface. It is an application that provides information for the web-browsing experience with an individually tailored, easy-to-use, repeatedly tested interface. Net Trust is focused on authenticating web sites *to individual users* instead of authenticating individual users *to web sites*. This paper describes the theory, motivation, and interaction behind Net Trust. Unlike commercial social browsing systems, Net Trust limits the distribution of identifiable personal histories to user-defined social networks. The goal is to enable users to control access to their personal information.

II. GOAL

Net Trust is a security technology designed to defeat masquerade attacks and to reduce the efficacy of Internet-based fraud by providing source authenticating information to the individual user. Masquerade attacks seek to lead the victim to believe that the perpetrating malicious entity is really an entity the victim trusts. Ironically, this type of attack is enabled by the *lack* of information on the Internet. A phishing attack is an example of this kind of attack. Phishing is difficult to prevent because it preys directly on the absence of source identification information online. For example, a user receives an email communication from a sender who purports to be the user's bank. The sender is actually running a malicious phishing site. The link embedded in the email leads the user to a site that is nearly identical to the legitimate site the user has visited in the past. This type of attack is very difficult for the user to detect. Simultaneously,

there is very little that the legitimate institution can do online to demonstrate that it is the original and not a masquerade site. Phishing, botnets[1] and other Internet-based fraud schemes continue to be a substantial and growing problem.[2] The Federal Trade Commission reported that in 2004, 53% of all fraud complaints were Internet related; as recently as five years ago, Internet-based fraud was a rare phenomenon. The Pew Internet & American Life Project has noted that 68% of Internet users surveyed were concerned about criminals obtaining their credit card information, while 84% were worried about the compromise of other personal data.[3] Net Trust identifies sites that identify themselves as popular banking sites as masquerades based both on the information from the FDIC and information from peers.

A second type of attack that could be decreased by Net Trust is a "zombie attack." A zombie attack occurs when a web site downloads malicious code or exploits browser vulnerabilities to create a zombie.[4] For example, a study by Microsoft using monkey spider browsers[5] (browsers which spider the web but act like humans) found 752 sites that subverted machines via browser vulnerabilities.[6] Net Trust is

---

[1] A botnet is a group of subverted machines all controlled by the same malicious hacker. For example, the hacker can use all the machines to send spam or to host phishing sites. The number of machines used by the hacker is directly related to the amount of spam that can be generated and having many hosts makes it significantly more difficult to shut down a phishing site.

[2] Federal Trade Commission: Protecting America's Consumers, "FTC Releases Top 10 Consumer Complaint Categories for 2004," *FTC.gov*, http://www.ftc.gov/opa/2005/02/top102005.htm (accessed October 17, 2007).

[3] Susannah Fox, "Trust and Privacy Online: Why Americans Want to Rewrite the Rules," Pew Internet & American Life Project, http://www.pewinternet.org/ PPF/r/19/report_display.asp (accessed October 17, 2007).

[4] A zombie is a machine that has come under the control of a remote party after the remote party subverted the machine. In Voodoo cults, the zombie is the body that is enslaved by a magician. On the Internet, the zombie is the computer that is controlled by a remote hacker. A group of zombies controlled by the same person is a botnet.

[5] A web spider climbs over all the links on the web, touching each one from a site, backtracking, and then going to another site and repeating the process. A web spider crawls the links in the World Wide Web as a physical spider crawls the links in a physical web.

[6] Y. Wang and others, "Automated Web Patrol with Strider HoneyMonkeys: Finding Web Sites That Exploit Browser Vulnerabilities" (symposium, Network and Distributed System Security, Internet Society, San Diego, CA, February 2, 2006) http://www.isoc.org/isoc/ conferences/ndss/06/proceedings/html/2006/papers.honeymonkeys.pdf.

designed to integrate a list of malicious sites and inform the user of the risks of the sites. It could even interrupt the connection with an informational warning.

III. MOTIVATION - NET TRUST IN ACTION

Net Trust is built on the personal observations described below combined with formal economic and information security theory. Keeping a network secure, much like keeping a neighborhood safe, requires the contribution of every member of a group. A potent observation is that as the network expands, more people become network operators. If telephone calls were switched by human operators today, every man, woman and child in America would have to be a telephone operator.[7] End user responsibility and control has increased even as dialing has gone from seven to ten digits and the proliferation of mobile devices have effectively ended universal directory service while vastly increasing the difficulty of interaction. The cell phone interface is much more complicated than a ten digit pad. Cell phones enable more complex functions at the cost of simplicity. Yet even with increased complexity, the contribution of naive users as opposed to trained operators has increased.

The individual computer owner has the most to lose when a computer is subverted. Attackers can take actions for which the computer owner could be held responsible, such as downloading copyrighted content or using the computer to implement other attacks. Individuals also bear the cost of spyware and malware, particularly when the malware enables financial fraud or identity theft.[8]

The economics of security advances a strong argument that the organizational incentives in a security mechanism must be aligned with the investment required for operation of that mechanism. Many security failures are a result of incentive misalignment.[9] The party at risk should be the party capable of making the most effective investment to mitigate that risk. Net Trust aligns incentives with

---

[7] There were women who performed switching on phone networks until the nineteen sixties in America. Even as late as nineteen eighty there were electro-mechanical switches in local loops.

[8] L. Jean Camp, *Economics of Identity Theft: Avoidance, Causes and Possible Cures* (New York: Springer-Verlag, 2007).

[9] Ross Anderson, "Why Information Security is Hard – An Economic Perspective," in *Proceedings of the 17th Annual Computer Security Applications Conference* (Washington, D.C.: IEEE Computer Society, 2001) http://www.acsac.org/2001/papers/110.pdf.

adoption in two ways. First, users control their own information rather than paying a dot com for a service. Second, Net Trust creates an open market for third parties supported by users or specialist groups by including lists of safe sites provided by Site Advisor, the FDIC, or any other software provider. Under the current system, third-party evaluators are paid by merchants and web site owners; this presents an inherent conflict of interest.

Information sharing is valuable because it creates incentives to mitigate risk. Even when individuals report inaccurately, the ability to identify and effectively deal with computer security risks increases.[10] In fact, being more informed and sharing information are correlated with additional security investments.[11] This finding suggests that Net Trust will not only be valuable in itself, but also may increase overall user security awareness and investment (future research with Net Trust includes experimenting with Net Trust users to test for the existence of complementary investment, e.g., securing home networks).

If Net Trust generates an increased awareness among individuals, it could also contribute to an increased awareness among firms that are seeking customers. Currently, investment in Internet security is arguably inadequate.[12] Like firms, individuals suffer immediate costs and future risks from a loss of information integrity. In the case of firms, a security incident is associated with immediate loss of value. A study of the capital market valuation of security incidents found that a firm can lose up to 2% of its market value within two days of a publicized incident.[13] Individuals also need mechanisms that allow them to share information effectively. When individual users produce

---

[10] L. A. Gordon, "An Economics Perspective on the Sharing of Information Related to Security Breaches: Concepts and Empirical Evidence" (workshop, The Economics of Information Security, Berkeley, CA, May 16–17, 2002) http://www.cpppe.umd.edu/Bookstore/Documents/EconomicPerspective_05.17.02.pdf.

[11] Esther Gal-Or and Anindya Ghose, "The Economic Consequences of Sharing Security Information," in *Economics of Information Security,* L. Jean Camp and Stephen Lewis, eds. (New York: Springer-Kluwer, 2004): 95–105.

[12] Lawrence A. Gordon and Martin Loeb, "The Economics of Information Security Investment," in *Economics of Information Security,* L. Jean Camp and Stephen Lewis, eds. (New York: Springer-Kluwer, 2004): 105–127.

[13] Allan Friedman and Alessandro Acquisiti, "Cost of Privacy Breaches" (workshop, The Economics of Information Security, Cambridge, MA, June 2–3, 2006) http://www.heinz.cmu.edu/~acquisti/papers/acquisti-friedman-telang-privacy-breaches.pdf.

information goods, as opposed to when firms produce information goods, the result is called peer production.[14]

In the traditional security paradigm, security decisions are often made for the user, without the user's knowledge, by the software on the user's computer. For example, ActiveX[15] and SSL[16] are two widely used programs and both use hierarchies to determine whether a web site will be trusted by the user's browser. These commercially-determined hierarchies give meaningful technical rights on the user's machine to those sites approved by the technical authorities. The extension of technical trust in these systems is made one of three ways. First, trust may be extended by default due to the default setting on the user's browser. Second, the extension of trust may result from the user "agreeing" by clicking a button on a pop-up window. Or, third, trust may be requested, again via a pop-up or a page on the requestor's site, through difficult to read, technically detailed information (not unlike an End User Licensing Agreement (EULA)).[17] The result is a lack of meaningful information to guide an individual's trust decisions on the network. In contrast, Net Trust is designed to allow users to make socially informed choices of their own.

In many cases, the security questions that are asked by centralized third parties are not relevant to the trust decision of the consumer at home. The third party may certify that an individual provided valid payment for cryptographic certification. The party might also confirm some identity claim. The consumer at home is more concerned with distinguishing between a reliable or reputable merchant and a recent e-

---

[14] Roger Dingledine, Nick Mathewson and Paul Syverson, "Reputation in Peer-to-Peer Anonymity Systems" (workshop, Annual Association of Computing Machinery, San Diego, CA, June 9–12, 2003) http://www2.sims.berkeley.edu/research/conferences/p2pecon/papers/s2-dingledine.pdf.

[15] ActiveX is used to allow web sites to run particular types of code to enable particular interactions based on a decision hierarchy that is grounded by Microsoft and the employer of the user.

[16] SSL, the secure sockets layer, is based on a hierarchy where providers of SSL certificates sell certificates that are then recognized by the user's browser. There is little or no information provided to the user about the decision by the certificate provider, or the implications of the level of certificate provided.

[17] An EULA is an End User Licensing Agreement. The only study of consumer use of EULAs found that users do not usually read EULAs, and when they read them, they do not understand them. See Nathaniel Good and others, "User Choices and Regret: Understanding Users' Decision Process about Consensually Acquired Spyware," *I/S: A Journal of Law and Policy for the Information Society* 2, no. 2 (2006): 283–344.

commerce entrant. The consumer may want to know if friends and family have had positive or negative experiences with the online merchant. An individual may vaguely remember a previous site and want to know if a present site is the one previously visited. Third parties cannot provide this information; they simply fail to provide socially meaningful information and thus do not effectively inform individual trust decisions.

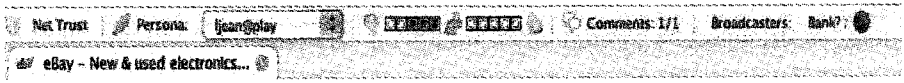
Net Trust combines information from personal browsing histories, social networks (as ratings) and third parties (as Boolean indicators). Combined, these three sets of indicators can both provide information to detect masquerade attacks and answer the questions about web sites that are important to consumers.

The Figures below show the ratings that may be presented to the user in the current version of Net Trust.[18]

Figure 1: Net Trust shows mixed ratings

Figure 2: Net Trust shows positive ratings

Figure 3: Net Trust shows negative ratings



Net Trust enables the integration of information from trusted parties, not simply information from unknown *third* parties as is the tradition in current security and cryptographic systems. The removal of the word "third" in the traditional trusted third-party construct indicates that the individual makes the final trust decision, not the "trusted" party. There is no root that determines which parties are

[18] The details of who provides the ratings and which mechanisms are used to create them are provided in the latter portions of the paper.

trusted. In trusted third-party systems, the browser manufacturer, employer or other third-party determine who is a trusted party. With Net Trust, users select trusted information providers.

The Net Trust user, not the distributor or developer, makes the final determination of which parties are trusted. This is the functional difference between SSL or ActiveX type security models, which the user is provided a mechanism to place trust in an unknown third party and Net Trust. Many of the mechanisms that enable trust and defeat fraud off-line are social mechanisms or physical mechanisms that cannot be re-created online.[19] In contrast, parties who engage in e-commerce are often separated geographically, temporally, and socially.[20] Consider the two companies in Figure 4 on the following page. Both are places where one might complete a transaction. If the two institutions were separated by meters, as opposed to an ocean, the differences would still be obvious. The Bank of Scotland has greater wealth, which indicates greater transactional security. The Bank of Scotland also has a building that reflects its history.

In economic terms, the Bank of Scotland is more trustworthy and is able to *signal* this quality through the construction of an impressive façade, its location at a prestigious address, and its highly ordered self-presentation. Compare this to The Check Cashing Place in Racine, Wisconsin. The façade of The Check Cashing Place indicates high competition, low overhead and a history no longer than the storefront lease. Online, none of the information that is so abundant in real life interactions is available. Geographical cues can be easily falsified but there are few expensive addresses in the infinitely expandable domain name system.

All domain names cost the same amount to register. Of the two below one institution holds millions in deposits. The other is trusted only to provide cash upon the presentation of a payment instrument. Even the range of payment instruments available at these facilities is not available on the Internet. Online, any payment requires transmission of information that can be reused by the receiving merchant for fraud, e.g. a credit card number. Any transaction on the Internet requires trust that would be worthy of the Bank of Scotland,

---

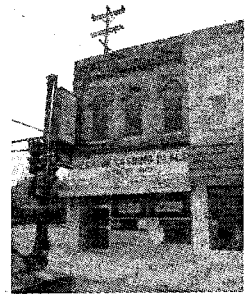
[19] Jens Riegelsberger and Angela Sasse, "Trustbuilders and Trustbusters: The Role of Trust Cues in Interfaces to e-Commerce Applications," in *IFIP Conference Proceedings* (Deventer, Neherlands: Kluwer, 2001): 17–30; Helen Nissenbaum, "Securing Trust Online: Wisdom or Oxymoron?" *Boston University Law Review* 81, no. 3 (2001): 635–664.

[20] Sonja Grabner-Kraeuter, "The Role of Consumers' Trust in Online-Shopping," *Journal of Business Ethics* 39 (2002): 43–50; Ravi Kalakota and Andrew B. Whinston, *Readings in Electronic Commerce: SPHIGS Software* (Boston: Addison-Wesley, 1996): 251–282.

but that information must be provided with less information about location, stability, and history than that presented by The Check Cashing Place.

Figure 4: Comparing financial institutions in the physical realm

Financial Services in Scotland, UK

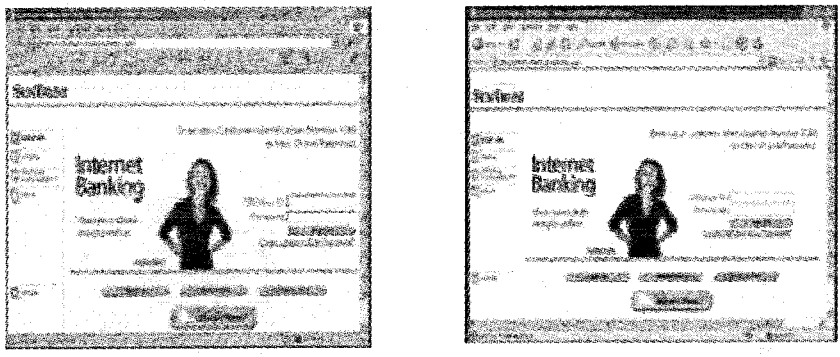


Financial Services in Racine, WI, US

Online, these virtual sites would be distinguished only by the web site design, domain name, and corresponding SSL certificates. Imagine that The Check Cashing Place renamed itself BankScotland. It could purchase any domain name, regardless of its registered business name. BankScotland and ScotlandBank, for example, were both available at the time this paper was written. An Internet user could easily confuse the two. In contrast, brick and mortar businesses can invest in physical infrastructure and expensive physical addresses to differentiate their prestige, customer service and reliability. To emphasize the point, consider the two pages below. The first image is provided to online customers by Sun Trust Bank. The second image was captured from a phishing attack that used a subverted computer at the business school at Columbia University controlled by a criminal entity (quite possibly on another continent).[21]

[21] Tyler Moore and Richard Clayton, "An Empirical Analysis of the Current State of Phishing Attack and Defense" (workshop, The Economics of Information Security, Pittsburgh PA, June 7–8, 2007).

Figure 5: Comparing putative financial institutions online



Sun Trust Sun Phish

The information infrastructure should have the most possible information about a source. However, it is easier to differentiate two types of stores (e.g., discount and prestige) offline than it is to distinguish between a bank and a criminal hideout online. This is because the possible sources of information, the number of channels of information, are greater offline than online. Online information is also easier to falsify. Even some things that don't seem notable off-line, such as background noise, odor, and number of other customers in a store cannot be verified or transmitted online. Thus the types, amount, and validity of information on the information infrastructure is more difficult to evaluate than the information online. In the figure above, the validity of the online information is difficult to judge. Security experts immediately recognize a false site as do many users, based on the absence of the icon indicating a SSL certificate. However, these cues can be falsified. My own web site (www.ljean.com) has as its icon a lock that indicates a secure site on Internet Explorer. But this is only a conceit, there is no corresponding security.[22]

The goal of Net Trust is to enable individuals to make better decisions about online resources by allowing individuals to use their own information, rather than forcing them to depend on the information provided by remote third parties. Net Trust makes these signals unique to each user and is more secure than easy to copy visual cues (e.g., the TRUSTe COPA seal or the Better Business Bureau seal)

[22] I was curious if this misdirection would work, and would continue to work with IE. It appears to at this time.

and more understandable than certificates. Net Trust is expandable and can be modified or used by other programmers without charge.[23]

IV. PEER SIGNALING IN RESOURCE ALLOCATION

The literature on peer production was popularized by open code and peer-to-peer music sharing systems.[24] Peer production has been found to have many advantages over firm-based production. Peer production changes the modularity, granularity and cost of integration of a produced good; it shifts the production costs to those most able and willing to bear them.

The peer production of information in Net Trust is highly modular. The granularity of Net Trust's implicit rating system is the URL. The use of social networks to group people affords Net Trust many advantages. Although limiting the radius of contacts from which information can be gleaned vastly constrains the total quantity of information, the smaller egocentric network of each individual is composed of individuals the user trusts.[25] Users will trust themselves not to invite a malicious actor into their personal networks and are less likely to trust the judgment of friends. Trust may be transitive, as a commonly cited model indicates, so long as it is within a finite radius.[26] A smaller network, both numerically and socially, has a smaller chance of containing a malicious node. A common computer science practice is to require individuals to create explicit numerical weights that reflect how much each node trusts their neighbors. The Net Trust model requires no such calculation; personal selection is a Boolean indication of trust.

A decrease in free riding is a possible outcome of a small network. Within a small social network, the incentive to cheat or free ride is less severe than in a larger anonymous system. Abusers of a small social

---

[23] Net Trust is distributed under a modified BSD license, meaning that the code is available if someone wants to improve it, alter it, or distribute it. All that is required is that credit is given to the original creators. Most code is distributed in a form that makes reuse impossible. For more details on licensing see Lawrence Lessig, *Code and Other Laws of Cyberspace* (New York: Basic Books, 2000).

[24] Yochai Benkler, "Coase's Penguin, or Linux and the Nature of the Firm," *Yale Law Journal* 112, no. 3 (2002): 435.

[25] These individuals are trusted in the social sense and not in the cryptographic provable sense.

[26] Thomas Beth, Malte Borcherding, and Birgit Klein, "Valuation of Trust in Open Networks," *Lecture Notes in Computer Science* 875 (1994): 3–18.

network will benefit less from unfriendly behavior because there are fewer known people to damage; (unlike with a centralized recommender), the incentives are further reduced because each member shares social ties with their immediate neighbors. The use of implicit data means that free riding, or having the system passively obtain information from others without generating any information, will still be an issue in Net Trust. When the alternative is non-use, free-riding should be encouraged.

There are practical advantages to a small network. The number of communication links scales as a square of the total network size; a smaller network decreases overall traffic.[27] Obtaining the same outcome with less traffic is more efficient. In addition to scaling up more easily, the fact that only a limited number of individuals are needed for the system to work means that Net Trust can grow without waiting for a critical mass. Finally, early usability tests indicate that individuals are more interested in signals from their immediate social network than in signals generated by global systems. Social networks may differ in terms of their perception of what is a legitimate or desirable site. For example, few who shop at Prada are likely to embrace Kmart's shoe sales; while few who buy shoes at Kmart will find Prada's pricing reasonable.

Quality does not need to be a global property, as long as it has a local meaning inside a given segment of the social network. Additionally, many commercial questions have local answers, none of which would be globally correct. "Where can I buy good shoes?" "Where is a good place to eat?" or "Who sells the best window treatments?" each have a multitude of locally correct answers, but only the answer that is specific to the location of the question's asker is desired. Other questions have only partisan answers. For example, conservative web sites might publish a range of truths and imaginative falsehoods about liberals, which would be accepted uncritically by some social networks and completely rejected by others. Similarly, there are a wide range of religious web sites and anti-religious web sites, both of which no authority could rank to the satisfaction of all people.

It has been demonstrated that the presence of a small, persistent application dedicated to a specific purpose will raise user awareness and consciousness about that concept, even if use of the particular

---

[27] Sergio Marti, Prasanna Ganesan, and Hector Garcia-Molina, "DHT Routing Using Social Links" (workshop, Peer-to-Peer Systems, San Diego, CA, February 26–27, 2004) http://dbpubs.stanford.edu:8090/pub/showDoc.Fulltext?lang=en&doc=2004-4&format=pdf&compression=&name=2004-4.pdf.

application is minimal.[28] This echoes the finding that suggests security information and investment are complements.[29]

The Net Trust system was designed first and foremost to address malicious, fraudulent and masquerade web sites. The basic model, however, is extensible. Net Trust incorporates implicit behavior-driven ratings, explicit individual recommendations and personally selected trusted parties. Conceptually, this user model would be useful for any resource of unknown quality when the resource quality is static. That is, a bad resource cannot strategically behave as a good resource some fraction of the time. For implicit information to work, the distribution of resources must not be independent of the distribution of users across the system. That is, for Net Trust to work, users in a self-selected social network should be more alike than a random group of strangers. In fact, the correlation between the web sites visited by participants in a social network does not have to be large if certain assumptions are made about the social network structure.

There are many situations in which the conditions listed in the previous paragraph (about bad actors and social networks) do hold. Sharing with trusted social contacts is superior to sharing with a group of strangers. With the inclusion of an implicit rating, social network-driven recommender systems could be used to address the additional generic quality problems. For example, interdisciplinary researchers cannot reliably ascertain which research journals are superior in fields that are not their disciplinary home. Net Trust can be used to track either the publishing or the reading habits of an academic peer group, allowing each member to gauge the relative importance of a journal based on its readership. Net Trust may add value beyond mere indication of phishing sites. That is, it may be enjoyable to use even by people who are never phished. That increased value will make Net Trust more usable since it will be perceived as worthwhile and more frequently used. Thus, Net Trust will provide more detailed information during browsing.

Peer production could enable production levels that cannot be achieved with centralized capital. The context information that can be

---

[28] Lorrie Faith Cranor, Manjula Arjula and Praveen Guduru, "Use of a P3P User Agent by Early Adopters," in *Proceedings of the ACM Workshop on Privacy in the Electronic Society* (New York, NY: ACM, 2002): 1–10.

[29] Esther Gal-Or and Anindya Ghose, "The Economic Incentives for Sharing Security Information," *Information Systems Research* 16, no. 2 (2005): 186–208.

re-embedded by constraining production to known, trusted and similar people in a social network is very powerful.

V. INTEGRATING PRIVACY-ENHANCED SIGNALING INTO BROWSING

The previous sections described the goals and motivation of Net Trust and presented a theoretical argument for obtaining information about peers to inform browsing. The actual mechanisms of Net Trust are described in this section.

Net Trust is a highly distributed, user-centered, usable trust management architecture that is resistant to spoofing, sybil attacks,[30] and web scripting.[31] Net Trust is a toolbar based interface. Eventually, Net Trust will support a p2p back end. Currently, Net Trust has a thick client and a thin server.[32] The client stores data, anonymizes data, calculates reputations from the social network, and manages the social network (e.g., confirming the relationship to an email address and a Net Trust identifier). The client authenticates data from third parties. User interaction is required to define the social network and make optional comments. Ratings are generated from observed user behavior, and then made pseudonymous for server-based distribution. The server authenticates this pseudonymous information when the data are written to the centralized server, then stores and distributes it.

Some toolbars target specific threats like phishing. Spoofguard is one of the toolbars that use the real-time characteristics of the phishing sites themselves; for example, links, images, lack of a SSL certificate, or misdirection in the links. Other toolbars are not designed to provide contextual or socially meaningful information. In other toolbars, the variables used to evaluate "trustworthiness" are under the control of

---

[30] A sybil attack occurs when one malicious party obtains many identifiers or accounts and then uses each to rate the others highly. For example, before eBay had a credit card requirement, it was possible to form hundreds of accounts, run fake auctions then each fake winner would rate the seller highly. After obtaining high ratings, real auctions were held by malicious but highly-rated merchants.

[31] Web scripting attacks occur when a web site appears to perform actions on behalf of the user. For example, a web site may include a command (e.g., script) to repeatedly search a term on Google, select a competitor's ad, and thus drive up the competitor's advertising cost to Google without providing the corresponding visitors. Obviously Google also has strong scripting detection; nonetheless attempts are made.

[32] Thick and thin refer to functionality. A thick client does many tasks and is complicated, while a thin server does little.

the malicious agent. In contrast, Net Trust uses features that are not under the control of the malicious agent: the user's social network, the user's browsing history and the browsing histories of the user's social network. In addition, the Net Trust toolbar takes advantage of a unique characteristic of phishing sites to prevent one phishing victim from misdirecting others: the temporal history of phishing sites. Phishing sites go up, are identified, and are taken down over a short period of time. Net Trust integrates the impermanence of phishing sites into its reputation system.

Net Trust appears to the user as a toolbar plug-in for a web browser. The toolbar is the main source of information to the user. The application has five primary components: the social network, the reputation system, the third parties, the interface, and the data distribution network. Recall Figures 1, 2, and 3. These illustrate how Net Trust in its current instantiation appears on various web sites: a technical gossip site, a bank, and an auction site with a reputation mechanism that privileges merchants over customers.

The following Figures provide a close-up of the two sources of information available for the individual. Figure 6 shows the peer ratings while Figures 7 and 8 show the only possible centralized ratings.

Figure 6: Peer ratings

Figure 7: A positive centralized rating

Figure 8: A negative centralized rating

As shown in Figure 6, Net Trust integrates social network information. Individuals may have multiple social networks: home, family, hobby, political, or religious. Regardless of the level of overlap, information from one social network may be inappropriate for another social network. Not only do people share different information with different people, but also different social networks

are associated with different levels of trust.[33] For example, professional colleagues can have much to offer in terms of the evaluation of professional web sites, but professional interactions are not characterized by the same level of openness as family; professional networks are not systematically used to request intimate or religious information. Because of these differences, overlapping contexts can cause a breach in privacy.

In order to support the construction of boundaries between a person's varying roles, the application allows a user to have multiple identities (e.g., pseudonyms) coupled with the individual's multiple social networks. Pseudonyms engage in different social networks and members of that social network are called "buddies." This is intended to indicate the similarity to other online connections and to indicate that the standard for inclusion may vary. "Buddy" is also sufficiently vague to communicate that there is no standard for strength of the network tie. The social network indicator as seen by the individual is shown in Figure 9.

Figure 9: ljean@work is distinct from ljean@play or ljean@home

When a user leaves or clicks out of a web site, the URL is associated with the pseudonyms visible in the toolbar. Choosing to associate a site upon departure instead of arrival allows users to make informed selections of web sites. Once a web site has been identified as associated with a pseudonym (in the figure shown, the pseudonym is "ljean@work"), the user no longer has to select that identity when visiting the associated web site. If ljean is in work mode, and then visits a site she has previously identified as associated with the ljean@home pseudonym, Net Trust will change pseudonyms at the site. Therefore, after a web site has been associated with a pseudonym, all future visits correspond to that pseudonym, regardless of the web site selection at the time of site entry. Thus, individuals have to make pseudonym choices only on new web sites. Presumably, individuals will select a default pseudonym. Then they can possibly select different pseudonyms for different machines, like those at their work or home.

---

[33] Judith Donath and Danah Boyd, "Public Displays of Connection," *BT Technology Journal* 22, no. 4 (2004): 71–82.

VI. THE BUDDY LIST

An essential component of this application is the re-embedding of a user's existing social network into his or her online browsing experience. In brick and mortar commerce, physical location is inherently associated with the social network, as exemplified by the corner store, or by regulars at businesses and local meeting places. Net Trust uses social networks to capture virtual locality information in a manner analogous to physical information. Net Trust implements social networks by requiring explicit interaction of the user. The Net Trust user creates a "buddy list" containing the social network associated with a pseudonym. Using the Net Trust invitation mechanism, a user sends a request to a buddy asking for authorization to add them to the user's buddy list.

Once the buddy approves the request, the user can place the buddy in the social network defined by the appropriate pseudonym. Social networks can be presented for user consideration from importing Internet Messaging (IM) lists, email sent lists, or pre-existing social network tools such as MySpace, Orkut, Friendster, Facebook or LinkedIn. Net Trust requires that the individual issuing the invitation to his or her buddy know the email or IM name of that buddy. The invitation includes file location information that must be integrated into the distributed file system. The following description identifies a file name as the minimal adequate locator.

Consider a Net Trust user named Alice who has an associate named Bob. Unlike standard cryptographic protocol descriptions, we assume that Bob and Alice have established a virtual social history. Before inviting anyone to a network, Alice creates a pseudonym. Once the pseudonym is created, she creates a set of asymmetric keys, both public and private. For simplicity, call the pseudonym Alice@work. The private key allows Alice to confirm that any message from Alice@work came from Alice@work to anyone with the corresponding public key. Alice sends an invitation with a nonce to Bob. A nonce prevents replay attacks and ensures freshness. The public key prevents anyone else from associating themselves with Alice's pseudonyms after the initial introduction. The public key is not published.

The example can only continue if Bob agrees to join the system. When Bob joins the system, Alice will share Alice@work's history with Bob's chosen pseudonym. The history-based reputation information is contained in a file or feed that is identified by a 128-bit random number. The feed or file will not include personally identifiable information. Since Alice initiated the invitation, she sends

Bob her file locator and a key used to sign the file. Then Bob will send his file locator and a key used to sign his feed. Part of Bob's choice includes filling out information about his affiliation with Alice, her name and his corresponding pseudonym, as well as a review date for her inclusion. Thus, interaction is designed to cause some cognitive dissonance when joining a stranger's network by demanding unknown information in order to consummate the introduction.

Indeed, social network sizes are fixed so that position in someone's social network has value. If social networks were expandable to the thousands, then choosing to join someone's network would be the default. Limiting the number of possible participants in a pseudonym is designed to decrease the likelihood that a stranger will be able to join. After distribution of Net Trust, we hope to implement experiments to test how likely Net Trust users are to accept a stranger to their social networks.

After this introduction, Alice and Bob update each other's own local reputation-based signals by sending out information. Alice and Bob update their own ratings by periodically downloading each other's published files. The files, designated "filename," include URLs, ratings and dates. Bob's ratings are then integrated into Alice's toolbar as Bob's opinions of sites. Using the data distribution system, Alice's client can read Bob's ratings and display them to Alice. Similarly, Bob's client can read Alice's ratings.[34]

In the proposed initial instantiation of Net Trust, different individuals' opinions are not to be weighed differently. Segregating individuals into social networks creates implicit weighting. Some systems assume that individuals should be provided with different trust weights because some contribute more than others.[35] In contrast, our system allows the user to evaluate his or her context based on the provision of the information. While the proverbial grandparent might not be as apt to discriminate between legitimate and malicious sites as a computer savvy co-worker, she may have extensive knowledge of hunger-based charities from volunteer work or detailed knowledge of travel locales from personal experience. Therefore, our initial design asserts that there is no single trust weight for an individual across all contexts.

---

[34] The importance of traffic analysis underscores the use of Tor in this system. Tor can prevent traffic analysis.

[35] Roger Dingledine, Nick Mathewson and Paul Syverson, "Reputation in P2P Anonymity Systems" (Workshop on Economics of p2p Systems, Berkeley, CA, June 5–6, 2003) http://freehaven.net/anonbib/cache/rep-anon.pdf.

Figure 10: Net Trust View

By simply hitting the icon of two people between the left, green, and right, red, rating bars, the user will see an enlarged view of their social network and pertinent browsing statistics. User-selected icons are displayed for ease of identification and personalization. Net Trust also allows for the addition of third parties who make assertions about trust as shown in the toolbar above. Centralized trusted parties provide these ratings. They are called "broadcasters" in this model to emphasize that they distribute but do not collect information.

While buddies share information by both sending information and obtaining regular updates, broadcasters only distribute information. Broadcasters use a certificate-based system to distribute their own files with Boolean ratings. Such lists of "good" and "bad" sites are sometimes called white and black lists or green and red lists. These lists are stored and searched locally to prevent the need for the Net Trust user to send queries that indicate their browsing habits. Requiring a web query for searching would create a record of the client's travels across the web, as with Page Rank records on the Google toolbar and the Microsoft anti-phishing toolbar. Broadcasters' ratings are shown as positive with a happy face, negative with a "yuck" face, and no opinion as blank.

Early user tests found that signals less blunt than smiling and "yucking" faces were confusing.[36] The default on a URL that is not included in the ratings provided by the broadcaster is to have nothing displayed. Net Trust users will be able to select their own broadcasters. To mitigate any possible harm, there is a maximum lifetime for any green list. Broadcasters can be removed but it is not possible for an attacker to replace one broadcaster with another even if the first one has been removed. Overriding that feature requires that an attacker have write permission on a user's drive. At that point, user trust of web sites becomes a negligible measure of security. Since the

---

[36] "Net Trust: A Cure for the Phishing Social Disorder," *Phishing* (Berlin, Germany: Springer, 2006).

broadcasters provide important information, like any other trust vector, subversion of that trust can cause harm. However, since broadcasters only inform trust decisions, the harm is limited. If a broadcaster provides bad information, the source can be detected by the user. Compare this with ActiveX or the addition of trusted certificate authorities which alter authorization and then access on the user's machine. With ActiveX or other certificate-based systems, malicious action from code cannot be observed during regular use by the user.

The security of this system depends on the ability of the user to identify network participants' reliability and to prevent leakage of the key used to share history. If attackers can rewrite histories, the system is a net loss in security terms. There is no universal identity infrastructure on which this system can depend. Invitations are issued by email, and email-based identity authentication is arguably tenuous. Social viruses have long utilized the lack of authentication in email to increase the likelihood of victims taking the necessary action to launch the virus. However, by requiring the response, this mechanism cannot be subverted by mere deception in the "from" field.

The security design was based more on economic rather than traditional security. The Net Trust system as modeled, which relies upon economics assumptions, will create value for users and increase the difficulty of certain types of financially motivated attacks. The next section describes the reputation system.

VII. THE REPUTATION SYSTEM

The current reputation system has been subject to extensive agent modeling and was constructed based upon the results of this system modeling. An initial visit will log the web site in to the system on the basis of domain names. This will create an initial rating of "one." That rating will decay uniformly so that if the site is not visited again the rating decreases to 0.5. Each time the web site is visited the rating will double to a maximum of n. Currently, n is set to ten. When a site is rated explicitly the explicit rating will remain without decay. Only explicit user action can change a rating based on previous explicit user action. In the case of a negative rating, the social network window shows a large red bar connecting the user to the site. The lowest implicit rating is zero.

For each web site that is associated with a pseudonym there are one of two possible records. There will either be an explicit rating and the URL, or the information required to calculate an implicit rating and a URL. Current phishing statistics suggest a value of delay time before an initial rating is entered of not less than twenty-four hours;

however, this may change over time. One system design question is whether users should be able to easily alter this delay time. Should the system be designed to allow an easy update if a new attack with a greater temporal signature is needed? If the value of the delay time is too low then attack sites could change victims to supporters too quickly. Thus, being able to increase the value offers the opportunity for a more secure mechanism. However, the value to altering the delay time can itself become a security risk because a phisher could convince users to set the delay time at zero.

To summarize the reputation system, a single visit yields the initial rating of one after some delay. The delay time prevents those who are phished early from becoming agents of infection and supporting future phishing attacks. Then, as the number of visits increases, the score itself increases in value. The lowest value for a visited site that has not been manually rated is zero. The greatest reputation value for any site is five. The least reputation value of any site is negative five. Negative ratings can only be generated by explicit user ratings, while positive user ratings are generated by individual browsing habits.

Consider a phishing web site which broadcasters label "bad" or "neutral." No member of the social network will have ever visited the site. While this may not deter someone from entering a site to shop for something unusual, it is an extremely unlikely outcome for a local bank, PayPal, or eBay. For example, it is not surprising if no one you know has visited www.piratemod.com/ unless one of your friends loves pirate gear. On the other hand, there should be positive ratings and history for PayPal or eBay. In order to increase the efficacy of the toolbar against phishing in particular, one element of the project entails bootstrapping all banking sites. Those web sites that are operated by FDIC-insured entities are identified by a positive signal (a smiley face). Those web sites that are not FDIC institutions are identified by a negative signal (a "yuck" face). The icons are shown and described in the previous section.

In addition, bootstrapping information can be provided by a compendium of shared bookmarks such as Give-A-Link or SiteAdvisor. PhishGuard generates a list of phishing sites and could be integrated into Net Trust. PhishGuard uses peer production of information by asking people to submit phishing sites but provides no privacy or other feedback. The FDIC, if it were the sole default broadcaster, would label any phishing site "not a bank." Without the inclusion of a FDIC listing, the Net Trust toolbar has a failure similar to many security mechanisms in that the user is forced to look for what is *not there*. Seals function only if they are noted both when they are present and *when* they are *missing*. The lock icon on SSL is replaced with a red icon but the user must notice that the lock is missing.

Providing only positive information and demanding that individual notice is missing is not as effective as providing both positive and negative information. However, few merchants and no criminals will provide negative information about themselves on a web site or in an email. In email, eBay messages include a header that indicates only messages with the eBay header are to be trusted. Obviously fake emails from criminals do not comply by including a flag to indicate that the expected eBay header is missing. Trust seals are easy to copy. It is apparent, valid economic signals for resource identification on the web are missing.

The long-term efficacy of the reputation system depends upon how similar social networks are in terms of browsing. Do friends visit the same web sites? Do coworkers visit the same web sites? For example, in the most general reputation system every user has a chance of seeing any one page. If there is a probability, p, that a user will correctly judge that a resource is bad, and the probability that the user the corresponding probability that the user would mislabel the source is 1-p, therefore a decision rule could be easily derived. This type of information is not currently available either for small social networks or available generally. If the user has a probability of p, to correctly judge a bad resource, and would mislabel it with the corresponding probability 1-p, a decision rule could be derived trivially. However, that information is not only unavailable for small social networks but also it is generally unavailable. For this type of research to be conducted Net Trust must be completed and have a group of users.

Using this reputation system with the assumption of different degrees of homophily,[37] the user modeling as described above indicates that Net Trust would provide a high degree of value in identification and annotation of web sites. Given that the ideal mechanism cannot be known because social network homophily is not known, the implementation is based on user modeling. Recall that user modeling indicates Net Trust will significantly increase the ability of end users to discriminate between resource types, e.g., good or bad. The model demonstrates that the inclusion of bootstrapping information will dramatically increase the ability of end users to discriminate accurately. Bootstrapping is providing by importing histories and adding third parties, i.e., the FDIC. The modeling of the reputation system indicates that the system - as proposed - will be

[37] Homophily means that people who are in a social network have similar browsing habits. Users are not uniformly distributed across the low-traffic sites on the web. Some sites of are interest only to a small population, such as members of a course at a university or members of one school or office.

valuable that assists users when they are distinguishing between safe and unsafe resources.

VIII. USABILITY STUDY RESULTS

Net Trust is only useful to the extent that it is usable. For this reason, Net Trust began with user testing. Twenty-five Indiana University graduate and undergraduate students participated in the first usability study of Net Trust, and fifty students participated in the second.[38] The students were from the School of Informatics. Initially, the participants of the usability study were asked to spend a few minutes investigating three web sites. The web sites were fabricated, especially for the purpose of a usability study, and controlled for content and interface.[39] The participants were asked to indicate if they would trust the sites with their personal identifiable information, including some financial information. In the first test, the toolbar was enabled in the browser and the participants were instructed to visit each of the three web sites again and complete one toolbar task on each site. The tasks included rating a site, adding and removing buddies, and switching between buddy and network view. The survey had been previously validated with two tests of undergraduates. For the Net Trust usability test, the toolbars were seeded with reputation information. In the second test, users were separated into those with and without the toolbars.

After examining the web sites, the participants were prompted to indicate their trust of the three web sites by taking into account the information provided by the toolbar. For the first two web sites, the toolbar showed a large number of "buddies" visiting the site, six out of ten for web site one, and eight out of ten for web site two, respectively, as well as positive or neutral ratings for the broadcasters. The last web site displayed only two out of ten friends visiting the site, and negative or neutral rating from the broadcasters. The toolbar significantly increased the propensity to trust a web site. The results demonstrate that the toolbar is successful in providing a signal of trust towards a web site. Even when the toolbar displayed a significant amount of

---

[38] Alla Genkina, "Re-Embedding Existing Social Networks into Online Experiences to Aid in Trust Assessment," (master's thesis, University of Indiana, School of Informatics).

[39] The similarity of the three web sites was tested at Loyola Marymount before the Net Trust experiments. L. Jean Camp, Cathleen McGrath, and Alla Genkina, "Security and Morality: A Tale of User Deceit" (lecture, Models of Trust for the Web, Edinburgh, Scotland, May 22, 2006) http://www.ljean.com/files/ WWW06camp.pdf.

negative ratings, as in web site three, the fact that a web site had been previously visited by members of a social network increased the propensity to trust. This finding is validated by my examination of trust mechanisms for which I argued that social networks are a powerful mechanism for enabling trust.

IX. PRIVACY CONSIDERATIONS AND ANONYMITY MODELS

The critical observation of the privacy that Net Trust – as it is proposed here – provides is the fact that the end user has control over his or her own information. Privacy can be violated when a user makes a poor decision to share information. However, the system does not concentrate data nor does it compel disclosure. There is a default pseudonym in the system that is shared with no other party (private) and another that collects no information at all (logout).

Net Trust is designed to ensure privacy in that the users can share selected information while controlling with whom they share information. This system shares web browsing information in a closed network of peers. In contrast, consider Furl and Del.icio.us. Both are designed to leverage the observation that each user has a unique view of the web informed by their own history and the history of others. In both systems there is significant centralized storage of user browsing and no explicit mechanism for user pseudonyms. Neither of these systems uses the developments in social network systems beyond simple collaborative filtering. Individuals do not select their own peer groups. As a result, information can be inappropriate and in some cases data are highly polarized. For example, a search for "George W. Bush" on Del.icio.us yields images of both the President and of various chimpanzees. Del.icio.us and Furl do have a commonality with Net Trust in that they both integrate peer-produced information.
Net Trust also differs from other social browsing mechanisms in that identity is not intended to be universal. There can be many "Bobs" so long as there is only one "Bob" in any particular social network. Effectively, identities are used as handles or buddy names to create a virtual implementation of a pre-existing social network. Identity construction assumes a previous context, so that meaning is derived from the name and context. Each person can construct as many pseudonyms as he or she desires, where each pseudonym corresponds to a distinct user-selected social network.

Net Trust is designed to have three default pseudonyms: user@home, user@work and private. Web sites visited under the private pseudonym are never distributed in the Net Trust data structures. There is an argument for keeping a "private" list stored.

The advantage is that users can inform their own personal browsing. The disadvantage is that the user may want nothing recorded. Our solution is to have a private pseudonym and an option to logout from the system. "Logout" is not considered a pseudonym. In all cases, if a user is logged in under a certain pseudonym, his or her web site activity will only be shared with the social network associated with that pseudonym and not with any other networks that might exist under different pseudonyms. The user may also edit by hand the list of sites that are shared with any particular social network.

Subsequently, a user's online activity is only used to inform the buddy view in other buddies' handpicked views. Becoming and offering oneself as a broadcaster requires downloading additional software and publishing a public key. The interface for broadcasters in our design accepts only single entry URLs and requires notations for each URL entered. Our design is directed at preventing anyone from becoming a broadcaster by accident. There is no implicit rating mechanism for broadcasters.

X. CONCLUSIONS

In general, privacy and security markets suffer from a lack of signaling. This decreases the overall demand for privacy and security information and products, as neither is necessarily trustworthy. One reason for the lack of reliable signaling is the economic incentive for producers of trusted information to sell that information. In an adaptation of Gresham's law, bad security can drive out good security when both demand the same price.[40] Indeed, high levels of Internet fraud make individuals less likely to interact with trustworthy sites. A daily usable version of Net Trust can be found linked at http://www.ljean.com/NetTrust/.[41]

---

[40] According to Encyclopædia Britannica, it is named for Sir Thomas Gresham (1519–1579), financial agent of Queen Elizabeth I, who was one of the first to elucidate it: "if two coins have the same face value but are made from metals of unequal value, the cheaper will tend to drive the other out of circulation; the more valuable coin will be hoarded or used for foreign exchange instead of for domestic transactions." Similarly, if there are cheap and unreliable security mechanisms which cannot be distinguished from expensive and reliable security mechanisms, all merchants will be pushed to adopt the cheap, unreliable security measures. Encyclopædia Britannica, online ed., s.v. "Gresham's law" (accessed November 20, 2007).

[41] Net Trust code for developed is currently available at http://code.google.com/.